ICT to visualize science and engage society

by Ross Mounce
Open Knowledge Foundation Community Coordinator, Open Science

A picture speaks a thousand words. This fundamental observation is important to note for the effective communication of narrative in most realms if not more so in science and its engagement of the wider society. Science is often complex, and modern 21st century science even more so, for example:

In the 1980s, I published a paper that presented seven hard-won data points showing the relationship between stress and velocity beneath a glacier. Two years ago, I was involved in an analogous experiment on the Antarctic ice sheet that created more than a billion times more data points. – Prof. Geoffrey Boulton, 2012

Back in the 1980s, pen & paper, chalk & blackboard, or even paint & canvas could easily visualize and communicate this scale of data. But how does one communicate a coherent visual narrative from over a billion data points? Any such visualisation of this scale absolutely *requires* computers and information technology solutions to convey such data-rich narrative.

The publication process of science is now completely digitised – no serious journal accepts handwritten manuscripts, or data on punch cards. Yet the dissemination of digital scientific publications has been relatively slow to take advantage of new technologies. Many journals online still retain many vestigial characteristics of their wood-pulp ancestors – restrictions on length, 2-column layouts, static black & white figures... all difficult to comprehend without an appreciation for tradition – this was the way things were done in the 20th century. The best journals at conveying narrative in a digital, visually-pleasing way are by no coincidence, new 'born-digital' journals like eLife, PeerJ, and PLoS ONE. They can accommodate interactive, dynamic, moving, colour figures – not just static 2D images. They supply their content in multiple formats, suitable for different devices & usages e.g. XML, HTML, mobile-optimized HTML & PDF. Most importantly, they are Open Access – anyone can view their contents from anywhere in the world via the Internet without having to pay-to-read, register, or login. Open Access also means that the figures and visualisations of the papers can be freely shared & remixed by readers as long as attribution to the original is given. In some cases this has clearly aided public engagement with examples ranging from neglected tropical disease research [1,2] to the evolution of shell pattern of *Conus* shells [3,4].

Yet I believe there are improvements still to be made. Science could look to the emerging field of 'data journalism' for inspiration with regards to the effective communication of data to the general public. The colourful interactive visualisation of huge & complex datasets by Guardian data journalists during the London Riots of August 2011 were both award-winning and captivating. The interactivity encourages online readers to dig into and fully explore all the nuances, subtly and variation in the data – in a way that the dry descriptive summary statistics, confidence intervals and p-values of traditional science just wouldn't engage. The use of colour, maps, networks and big data all aided a data-driven analyses of what actually happened in the riots and what caused them. It helped disprove the otherwise intuitive stories being written, with well evidenced narrative supported and emboldened by real data rather than just punditry or opinion. The scale of data analysed was immense. Millions of tweets for many analyses. No single tweet would do the story justice, but an analysis of all makes a compelling read.

The downside to bright, colourful, interactive infographics is that their beauty of presentation can be used selectively to either misconstrue or obscure important data. Science will need to take care to

balance the needs of engaging presentation with that of scientific rigor and statistical standards to uphold but the two need not be diametrically opposed.

A final thing that sets data journalism apart from most science is the free access to the underlying data. In order to make the journalism independently verifiable, where possible data journalists often provide the full data sets that they have analysed as part of the article. Not only free access, but it's also provided in immediately re-usable formats that are actually of use to others, with adequate metadata that can be readily interpreted. One rarely finds this in science – most supporting data is often carelessly buried in supplementary appendices inside PDF files with often cryptic metadata – completely unusable. Much could be simply done to improve this situation, and recent initiatives like FigShare, Dryad and LabArchives are helping to provide usable digital, sustainable, free access data stores for scientific data, but their uptake is far from universal in the scientific research community.

The future's bright for visualisation & narrative in science, but academics are far from leading in this area. Collaboration and knowledge exchange with data journalists and data vizualisation experts would no doubt help the future of public engagement with science.

References:

1. Krauth, S. J., Coulibaly, J. T., Knopp, S., Traoré, M., N'Goran, E. K., and Utzinger, J. 2012. An In-Depth analysis of a piece of shit: Distribution of schistosoma mansoni and hookworm eggs in human stool. PLoS Negl Trop Dis 6:e1969+. http://dx.doi.org/10.1371/journal.pntd.0001969

2. Gonzalez, R. T. 2012. This scientific article about poop may have the best title in the history of peer-reviewed research (possibly NSFW figures) *io9.com* Retrieved from: http://io9.com/5972925/this-scientific-article-about-poop-may-have-the-best-title-in-the-history-of-peer+reviewed-research-possibly-nsfw-figures

3. Gong, Z., Matzke, N. J., Ermentrout, B., Song, D., Vendetti, J. E., Slatkin, M., and Oster, G. 2012. Evolution of patterns on conus shells. Proceedings of the National Academy of Sciences 109:1367 http://dx.doi.org/10.1073/pnas.1119859109

4. Keim, B. 2012. The best scientific figures of 2012. Wired Magazine. Retrieved from: http://www.wired.com/wiredscience/2012/12/science-figures-2012/?pid=5585